Study on human Action Recognition Algorithms in videos

He Chun-Lin1 ,a Pan Wei2，b
1,2Computer College,China West Normal University, Nanchong, 637002,China
achunlin_he@163.com,achunlinhe123@126.com, b2645831@qq.com

Abstract

Human action recognition algorithms are the one of the effective schemes to analyze the action of the people in the surveillance environment. It is one part of the event recognition algorithm. There are many types of action recognition algorithms, and in this paper, some classical methods are surveyed. The detail procedure of human action recognition is described, and the basic theory application of some algorithms is also analyzed. Finally, the conclusion of this paper is given.

Keywords: Action Recognition; Support Vector Machine; Bayesian Network

Introduction

With the rapid development of information technique, especially there are more and more intelligent terminal devices, the images and the videos become more and more. There is more information in the image or in the video. For example, people focus on the human action in the still image or in the video. If the action can be recognized in these scenes, the model of the conditional surveillance can be changed. People can know the detail information of the human. Therefore, the human action algorithms can be classified into two types (figure 1), one is based on still images [1,2]. The other is based on videos. The algorithms based on still image have smaller information than the algorithms based on video. That is the reason that the video not only has the space information, but also has the temporal information. Therefore, most of these algorithms are based on video. In this paper, the algorithms are also based on video.

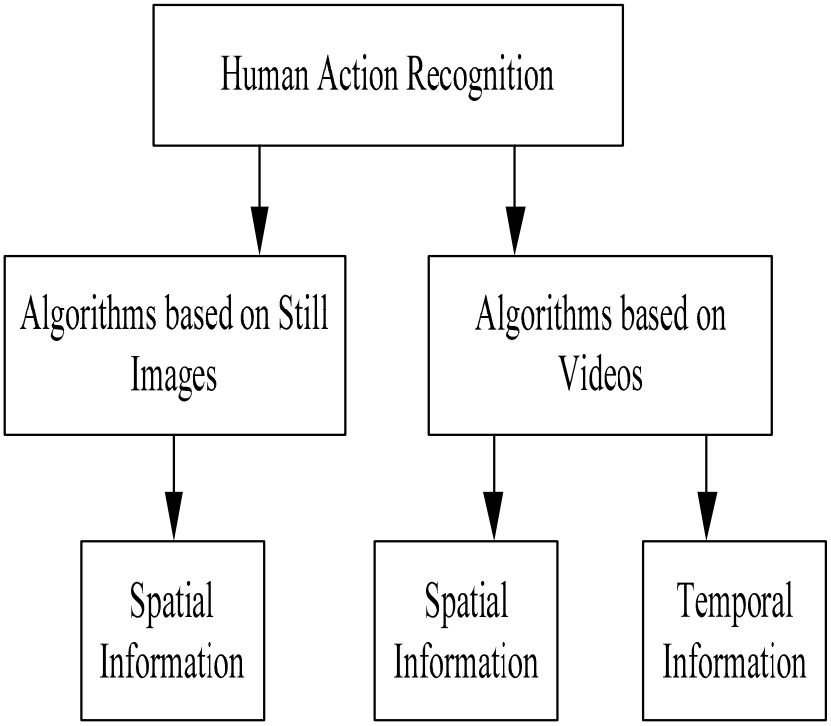

Fig. 1 The human action recognition's classification

The remainder of this paper is organized as the follows. In section II, the basic procedure of human action recognition algorithms is given, and then some algorithms of human action recognition are described in section III. Some advice of these algorithms and a conclusion is presented in the next section.

Basic procedure of human action recognition

Action recognition system includes following parts: pre-processing, the moving target detection, object classification, human body tracking, feature extraction, action classification, as shown in figure 2. When we get a video, the pre-processing is to clean the video by reducing the noise, sometimes; this step is included into the next steps. For human action recognition, moving objects detection is to extract the human body by examining the change region from the background image in a sequence of video frames. After the object detection algorithms, the objects are all detected, so the object classification should be used to extract the human body. There are methods for object classification, for example, the movement characteristics classification and shape classification. There are many tracking methods, they are: background subtraction method, optical flow method, block matching, the time difference method, active contour

models method, etc. In human body tracking method, there are mainly the models, for example, the method based on region, based on active contour, etc. Feature extraction methods can be divided into two categories, one kind is low-level image information method, and the other is the method based on high-level human body structure. Action recognition classification method mainly can be divided into state space method, the template matching method and the semantic description method.

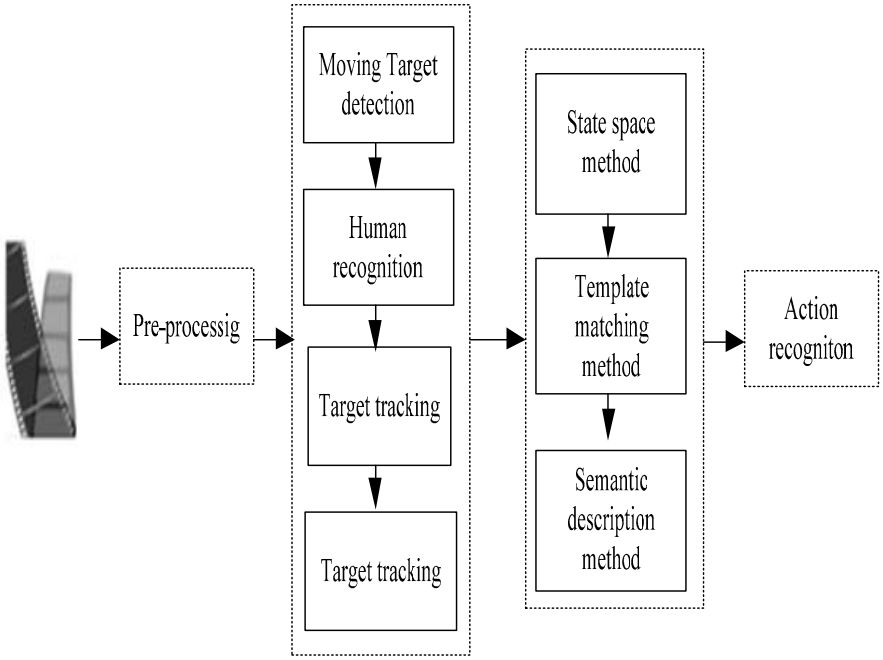

Fig. 2 Basic procedure in human action recognition's algorithms

The basic action feature extraction

The action feature extraction and recognition is the fundamental problem of the human action recognition. Selecting proper action characteristics is the first step in order to express the action characteristics as of the processed data as the corresponding characteristic data. And then enter data into the system action recognition to identify categories. The extraction of action characteristics means extracts the characteristics data. And the data can be reasonably representing the action from the video sequences of human action. Depending on the action of classes and scenes, usually choose different characteristics to represent the action of the target.

In the distant view video, the moving object trajectory should be extracted as action characteristics; while in the close-range video, for conduct human gesture recognition will need for human limb 2D or 3D modeling and then extract features

from the model. Feature extraction method in accordance with the nature of the action characteristics can be divided into two categories: one is a method based on low-level frame information[3-5], and the other is a method based on high-level human anatomy[6-7]. Low-level information always includes fowllowing aspects: target, movement speed, movement, the outline of the local space-time interest points, optical flow and characteristics of time and space, etc. For the simple character, the methods based on low-level information of features are widelly used, but the performance is worse than the high level information. High-level human body structure information refers to the structure presented by the human body posture; it can more finely describe human action than the low-level information. According to the use of different mannequin, such algorithms can be divided into three types: the method based on human point model, the method based on two-dimensional human model and three-dimensional human model. However, the approach based on high-level human structure typically has many parameters, training complex and has large amount of calculation. For example the high level of the action recognition include the method based on human point model, the method based on 2D human model, and the method based on 3Dhuamn model, the method based on the descriptive characteristics. The algorithms based 3D human model is the future direction of the action recognition.

The basic action recognition method

There are many methods of action recognition, and these algorithms can be divided into two classes: the method based on state space and the method based on template matching, the first one includes dynamic Bayesian networks, hidden Markova model, conditional random field, and so on. The other scheme includes frame to frame matching method and fusion matching method. And performance of the first method is better than the second method. However, the complex of the first method is higher than the second one (Table 1).

Table 1 Two Schemes comparing

method	schemes	performance	complex
Template matching method	dynamic Bayesian networks, hidden Markova model, conditional random field	better	higer
State space method	frame to frame matching method and fusion matching method	good	not high

Dynamic Bayesian network[8-10] is essentially a Bayesian network, having the same structure with Bayesian networks on the timeline, is a description of a stochastic process of change over time. In practical application problems, probability of each node is predicted through the DBN inference algorithm, and then you can get the joint probability of the event, and select the category with the highest probability to identify the classification. The algorithms based on hidden Markova model algorithm are not only has good results in processing the time-varying signal, but also has training and learning mechanisms and good ability to identify, and is one of the most widely used method.

Action recognition based on HMM algorithm[11] mainly consists of two parts, one is the stage of training and learning, given the observation sequence and initial HMM model parameters , new HMM model parameters can be obtained through training and the new HMM model parameters can be made to maximize the probability of classification, ie, to achieve better and more accurate recognition results; another is the identification and classification stage, given an behavior characteristic sequence of unknown categories, the probabilities of the new HMM model parameters belonging to different categories is calculated with the new model parameters and then find the category with the highest probability as the observed sequence's behavior.

Conditional random field(CRF)[12], this is a discriminant undirected graph probability learning model, is applied to label and segment ordered data. CRF is essentially a discriminant markov model, training CRF parameters by maximizing the classifier discriminant weights, rather than the joint probability of training sample. Not only can modeling the features of arbitrary observation sequence, but can adapt to overlapping features. CRF don’t need the assumption of conditional independence for observe characteristics, so these algorithms can better apaptive to the real-world condition. CRF is appropriate for dynamic time series modeling, can better recognize the simple actions, continuous and complex interaction behaviors.

Template matching method[13] is to compare testing video and the template video, find the template sequence which has the greatest similarity with the test sequence, then the category of the corresponding template feature sequence is the category of the test sequence. Template matching methods generally can be divided into frame to frame matching method and integration matching method. The first one is to compare the test sequence with template characteristic sequence frame by frame and calculate the similarity of both.State space approach is to define each static posture in the sequence of behavior as a state, and these states are linked to each other by the probability, any kind of motion sequence can be regarded as a traversal process between the different status at a time, joint probability is calculated as the criteria for the classification of behavior.

Conclusion

Action recognition based on video has been an important research direction In this paper, according to recent research, this paper give the classification of the human action recognition, describe the basic procedure of the human action

recognition, summarizes the basic methods of action recognition, including two aspects: feature extraction and the action recognition method. The first one includes the high level feature extraction and the high level feature extraction. In the action recognition method, we briefly describe the basic schemes based on dynamic Bayesian networks, hidden Markova model, and conditional random field. The future direction of the human action recognition is algorithms based 3D model.

References

[1] Thurau C, Hlavác V. Pose primitive based human action recognition in videos or still images[C]//Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE, 2008: 1-8.

[2]Poppe R. A survey on vision-based human action recognition[J]. Image and vision computing, 2010, 28(6): 976-990.

[3]Cheng Z, Qin L, Ye Y, et al. Human daily action analysis with multi-view and color-depth data[C]//Computer Vision–ECCV 2012. Workshops and Demonstrations. Springer Berlin Heidelberg, 2012: 52-61.

[4] Yilmaz A, Shah M. Matching actions in presence of camera motion[J]. Computer vision and image understanding, 2006, 104(2): 221-231.

[5] Zhu G, Xu C, Huang Q, et al. Action recognition in broadcast tennis video[C]//Pattern Recognition, 2006. ICPR 2006. 18th International Conference on. IEEE, 2006, 1: 251-254.

[6] Liu J, Kuipers B, Savarese S. Recognizing human actions by attributes[C]//Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011: 3337-3344.

[7] Ben-Arie J, Wang Z, Pandit P, et al. Human activity recognition using multidimensional indexing[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2002, 24(8): 1091-1104.

[8] Muncaster J, Ma Y. Activity recognition using dynamic Bayesian networks with automatic state selection[C]//Motion and Video Computing, 2007. WMVC'07. IEEE Workshop on. IEEE, 2007: 30-30.

[9] Park S, Aggarwal J K. Recognition of two-person interactions using a hierarchical Bayesian network[C]//First ACM SIGMM international workshop on Video surveillance. ACM, 2003: 65-76.

[10] Du Y, Chen F, Xu W, et al. Interacting activity recognition using hierarchical durational-state dynamic Bayesian network[M]//Advances in Multimedia Information Processing-PCM 2006. Springer Berlin Heidelberg, 2006: 185-192.

[11]Brand M, Oliver N, Pentland A. Coupled hidden Markov models for complex action recognition[C]//Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on. IEEE, 1997: 994-999.

[12]Wang S B, Quattoni A, Morency L, et al. Hidden conditional random fields for gesture recognition[C]//Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on. IEEE, 2006, 2: 1521-1527.

[13]Bobick A F, Davis J W. The recognition of human movement using temporal templates[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2001, 23(3): 257-267.